https://doi.org/10.31436/iiumej.v27i3.4388

A Modular, 3D-Printed, and Cost-Effective Near-Infrared Spectrometer with Machine Learning for Commodity Plastic Identification

WEI JIM TAN, ZULKIFLI MOHAMAD ARIFF*, ARJULIZAN RUSLI, MUHAMAD SHARAN MUSA

School of Materials and Mineral Resources Engineering, Universiti Sains Malaysia, Nibong Tebal, Malaysia.

**Corresponding author: zulariff@usm.my*

(Received: 6 May 2026; Accepted: 14 July 2026; Published online: 11 September 2026)

ABSTRACT: Efficient segregation of post-consumer plastics remains a critical challenge in achieving a sustainable circular economy because current sorting technologies are limited in cost, adaptability, and accessibility for community-level use. This study presents the development of a modular, cost-effective, and 3D-printed transmission near-infrared (NIR) spectrometer that operates in the 850 nm to 1100 nm range. The spectrometer was developed using open-source tools and resources such as Theremino and Scikit-learn, as well as commercially available optical components such as a desktop webcam, a diffraction grating, and an incandescent penlight. Validation of the spectrometer revealed minor wavelength deviations between theoretical and measured values. Measured values deviated by 5.1% with respect to the 546 nm mercury characteristic peak and 7.6% with respect to the 436 nm characteristic mercury peak emitted from a fluorescent lamp. This deviation is due to the low-cost optical components, but it did not impair the classification performance of the polymers. 504 plastic samples from seven material classes, namely acrylonitrile butadiene styrene (ABS), low-density polyethylene (LDPE), polyethylene terephthalate (PET), polymethyl methacrylate (PMMA), polypropylene (PP), green polystyrene (GPS), and red polystyrene (RPS), together with an empty baseline reference (MT) sample, were collected and processed using Principal Component Analysis (PCA), where 12 principal components were retained to capture 95% of the cumulative spectral variance. Classification performance was subsequently evaluated using Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), and K-Nearest Neighbors (KNN) models. After hyperparameter tuning, both the SVM and KNN models achieved the highest held-out test accuracy (82.24%). SVM was the most reliable overall classifier, with slightly better class-balanced performance (F1-score of 82.36%) than KNN (81.98%) and LDA (73.71%). The empty baseline reference spectrum was classified perfectly with a precision, recall, and F1-score of 100%, confirming the consistency of the baseline penlight source spectrum. RPS showed greater feature overlap with GPS, and LDPE with PP. This study demonstrated that a low-cost, transmission NIR spectrometer operating between 850 nm and 1100 nm, combined with machine learning, can accurately identify commodity plastics, offering a promising community-level solution.

ABSTRAK: Pengasingan plastik pasca guna secara berkesan masih merupakan cabaran utama dalam mencapai ekonomi kitaran lestari. Kekurangan teknologi pengasingan plastik sedia ada dari segi kos, kebolehsuaian, dan kebolehcapaian gunaan di peringkat komuniti telah menjadi penyumbang kepada cabaran tersebut. Kajian ini menumpukan kepada pembangunan spektrometer inframerah dekat (NIR) jenis transmisi melalui pencetakan 3D bersifat modular, kos efektif, dan mempunyai julat operasi dari 850 nm hingga 1100 nm. Spektrometer ini dibina menggunakan platform sumber terbuka seperti Theremino dan Scikit-learn. Selain itu, komponen optik yang mudah diperoleh seperti kamera web desktop yang diubah suai, kisi

pembelauan, dan lampu pijar jenis pen turut digunakan. Semasa proses pengesahan fungsi spektrometer ini, sisihan panjang gelombang kecil antara nilai teori dan nilai yang diukur telah dikesan. Berdasarkan dapatan pengukuran, terdapat sisihan sebanyak 5.1% bagi puncak ciri merkuri pada panjang gelombang 546 nm dan 7.6% bagi puncak ciri merkuri pada panjang gelombang 436 nm yang dipancarkan oleh lampu pendarfluor. Penggunaan komponen optik kos rendah mengakibatkan berlakunya sisihan tersebut, tetapi keupayaan spektrometer ini dalam pengelasan polimer tidak terjejas. Sebanyak 504 sampel plastik daripada tujuh kelas bahan iaitu akrilonitril butadiena stirena (ABS), polietilena berketumpatan rendah (LDPE), polietilena tereftalat (PET), polimetil metakrilat (PMMA), polipropilena (PP), polistirena hijau (GPS), dan polistirena merah (RPS), bersama satu garis dasar rujukan kosong (MT) telah dikumpulkan dan diproses menggunakan Analisis Komponen Utama (PCA), di mana 12 komponen utama dikekalkan bagi mewakili 95% varians spektrum kumulatif. Prestasi pengelasan kemudiannya dinilai menggunakan model Mesin Vektor Sokongan (SVM), Analisis Diskriminan Linear (LDA), dan K-Jiran Terdekat (KNN). Selepas penalaan hiperparameter, model SVM dan KNN mencapai ketepatan ujian ketahanan tertinggi sebanyak 82.24%. SVM dikenal pasti sebagai pengelas keseluruhan yang paling boleh percaya dengan skor F1 sebanyak 82.36%, berbanding KNN (81.98%) dan LDA (73.71%). Spektrum sampel MT berjaya diklasifikasi dengan sempurna dengan nilai ketepatan, penarikan semula, dan skor F1 sebanyak 100%, sekaligus mengesahkan konsistensi spektrum asas yang dihasilkan daripada cahaya lampu pijar. Namun, bahan seperti RPS menunjukkan pertindihan ciri spektrum yang lebih tinggi dengan GPS, manakala spektrum LDPE pula bertindih dengan spektrum PP. Kajian ini berjaya membuktikan bahawa spektrometer NIR transmisi kos rendah yang dihasilkan melalui pencetakan 3D dan pembelajaran mesin dapat beroperasi pada julat 850 nm hingga 1100 nm dan mampu mengenal pasti plastik komoditi dengan tepat di samping menawarkan penyelesaian terbaik pada peringkat komuniti.

KEYWORDS: *Near-Infrared Spectroscopy (NIR), Plastic Identification, Low-Cost Spectrometer, Machine Learning, Principal Component Analysis (PCA).*

1. INTRODUCTION

Global plastic consumption continues to rise because of its remarkable combination of properties and characteristics, making it essential to daily life; yet only 9% of all waste plastic generated worldwide is recycled. The most preferred forms of plastic waste management remain landfilling, which accounts for almost 50% of plastic waste, followed by incineration at approximately 19% [1]. Furthermore, the complexity of plastic polymers and their additives further complicates recycling because plastic products are often compounded, with polymer molecules that are no longer uniform and include copolymers, blends, and composites [2]. Most recycling systems rely on centralized post-sorting technologies such as Material Recovery Facilities (MRFs), where contamination is already significant and sorting accuracy depends heavily on expensive optical systems or trained personnel [1], [3]. Fourier transform infrared spectroscopy (FTIR) and other laboratory-grade instruments provide excellent identification performance but remain inaccessible for community or household applications because of their cost, required expertise, and limited adaptability. A key advantage of FTIR is its non-destructive measurement mode, which preserves material integrity and supports material recovery during recycling processes.

In this context, a promising alternative, near-infrared (NIR) spectroscopy, offers similarly non-destructive and rapid identification capabilities. NIR spectroscopy is widely used for polymer analysis because of characteristic absorption features associated with molecular vibrations in the wavelength range of approximately 800–2500 nm [4]. However, despite these

advantages, commercially available NIR spectrometers remain relatively expensive and are not readily accessible for community or household applications.

Industrial-grade NIR and hyperspectral imaging (HSI) systems are effective for high-speed material recovery, but their high investment costs range from $15,000 to $50,000, which renders them inaccessible for community-level applications. This study demonstrates that by utilizing 3D printing and consumer-grade silicon sensors, the cost of NIR technology can be reduced to a few hundred dollars, making it a viable tool for decentralized recycling initiatives, fablabs, and small-scale workshops [5].

Nonetheless, recent advancements in open-source hardware, 3D printing, and low-cost optical components could extend access to commodity plastic segregation capabilities for the general public. Open-source devices such as the RM90 Theremino spectrometer show that affordable designs can achieve competitive performance in the visible region comparable to a commercial USB4000 spectrometer costing over RM5,000, suggesting similar approaches may apply in the NIR domain [6]. NIR spectrometer instrumentation requires appropriate diffraction optics, guided by the grating equation and long-pass filters to minimize order overlap.

Several DIY spectrometers are available online that employ different spectroscopic methods. Two projects designed specifically for plastic identification are the “Plastic Scanner” by Straller & Gessler [7] and “Ozirma” by Vivien Henry (https://hackaday.io/project/179969-ozirma-near-infrared-spectrometer). The Plastic Scanner is highly compact and features an array of IR LEDs spanning 850 to 1650 nm paired with an InGaAs sensor, but these components are expensive and difficult to source. Ozirma is larger and utilizes a concave mirror, a rotatable reflection diffraction grating controlled by a stepper motor, and a single-pixel InGaAs photodiode. Additionally, DIY spectrometers not specifically developed for plastic identification offer useful design insights, such as the OSMS visible-range spectrometer for liquid analysis and a 3D-printable modular smartphone visible spectrophotometer [8].

Despite existing advancements in DIY spectroscopy, a significant gap remains between high-cost industrial systems and current open-source tools. Innovative projects such as the 'Plastic Scanner' have demonstrated the efficacy of using discrete LED wavelengths for material identification; however, extending these systems to longer NIR wavelengths typically requires InGaAs sensors, which can be expensive and difficult for the general public to source. In this study, we propose a complementary approach that utilizes more accessible silicon-based sensors. Although these sensors restrict the detection range to 850 nm to 1100 nm, our prototype uses continuous-spectrum measurements to extract a higher density of meaningful classification patterns from this narrower range. This provides a cost-effective alternative for decentralized, community-level plastic segregation without expensive longer-wavelength components.

The prototype features a modular, 3D-printed spectrometer that utilizes unconventional, low-cost materials such as razor blades to create sharp aperture slit edges and modified incandescent penlights to achieve continuous NIR acquisition. This hardware innovation improves accessibility by showing that standard silicon-based webcam sensors, traditionally limited to the visible spectrum, can be repurposed for NIR polymer identification up to 1100 nm through simple filter modification. These hardware capabilities are supported by an optimized chemometric workflow featuring a rigorous machine learning pipeline that uses Parallel Analysis (PA) to objectively retain principal components. This approach ensures a high classification accuracy of 82.24% for seven commodity plastic classes, demonstrating the system's robustness even when accounting for inherent hardware-induced spectral deviations.

https://doi.org/10.31436/iiumej.v27i3.4388

2. MATERIALS AND METHODS

2.1. Spectrometer Prototype Fundamental Components

The developed spectrometer, as shown in Fig. 1, was designed based on the principle of transmission spectroscopy, in which the plastic sample was positioned between the illumination source and the detector. The spectrometer design is governed by the grating equation Eq. (1), which defines the relationship between the groove spacing (d), the angle of incidence (θ), and the angle of diffraction (φ) for a given wavelength:

$$n\lambda = d(sin\theta + sin\varphi) \quad (1)$$

where n is the diffraction order. By utilizing a transmission grating with 1000 lines/mm (d=1000 nm), the system was optimized to disperse incident light across the webcam sensor. Spectral acquisition was performed using a desktop webcam (CW834M) as the detector after removal of its built-in infrared (IR) blocking filter. Because the system uses a silicon-based image sensor, the practical detection range was limited to approximately 1100 nm. An educational-grade transmission diffraction grating with a groove density of 1000 lines/mm was used to disperse incident light, while a Zomei 850 nm long-pass optical filter was incorporated to minimize overlap between diffraction orders in the visible and NIR regions. A modified Fable incandescent penlight served as the broadband light source and was powered using a regulated 3.3 V supply from an Arduino to reduce baseline spectral drift caused by decreasing battery discharge. The aperture slit was constructed using razor blades to produce sharp and straight slit edges for improved resolution.

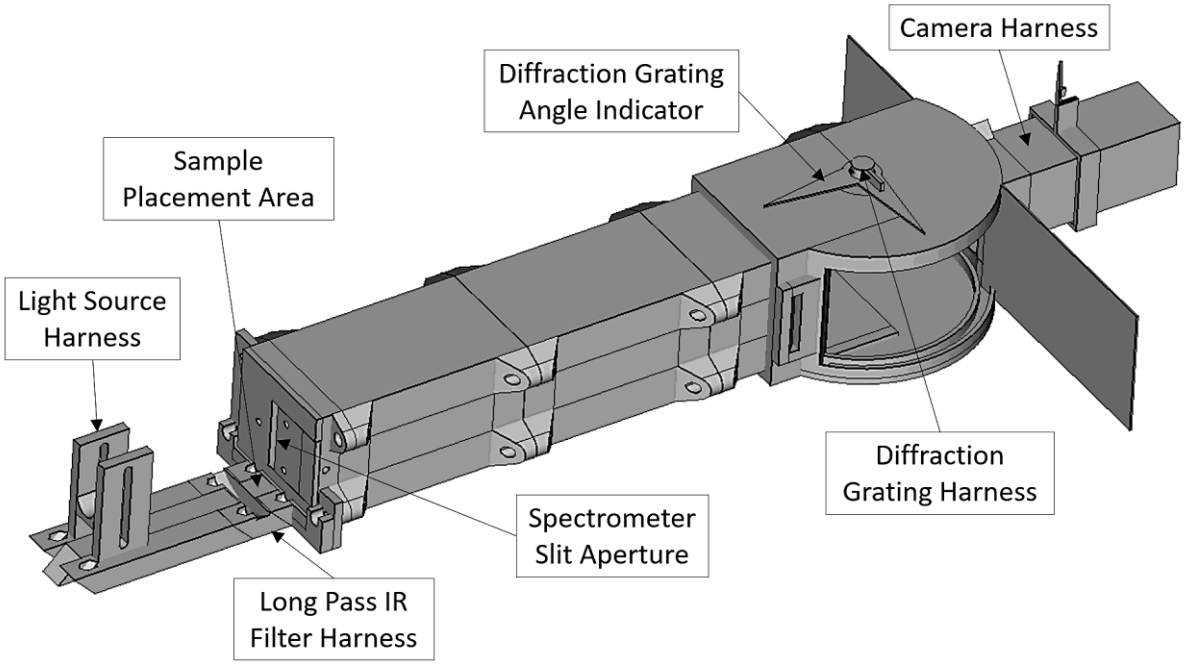

Figure 1. CAD 3D model of spectrometer

A total of seven polymer samples and one empty baseline reference sample were used in this study, namely acrylonitrile butadiene styrene (ABS), low-density polyethylene (LDPE), polyethylene terephthalate (PET), polymethyl methacrylate (PMMA), polypropylene (PP), green polystyrene (GPS), and red polystyrene (RPS), together with an empty baseline reference (MT) representing the absence of a sample. The GPS and RPS samples used the same base polystyrene polymer but differed in colorant. To ensure spectral consistency during model

development, all laboratory-prepared samples were compressed to a standardized thickness of 1.0 mm prior to measurement.

2.2. Spectrometer Operation and Data Analysis

2.2.1. Spectrometer Prototype Setup and Software Settings Standardization

The open-source Theremino Spectrometer Software was utilized. To optimize signal quality, the contrast setting was standardized at the maximum value to eliminate peripheral background noise inherent to the low-cost camera. The aperture width was standardized at 2.5 mm. The angle of incidence was set to 42.5° and the angle of diffraction to 22° to maximize NIR signal detection within the camera's sensor limits. These angles were determined by iteratively adjusting the diffraction grating and camera position relative to the optical axis. This configuration was finalized once the continuous spectral distribution was fully projected across the active area of the silicon sensor, validated through real-time visualization in the Theremino software interface.

2.2.2. Machine Learning Workflow

The machine learning workflow starts with a dataset of 504 samples and 460 spectral wavelength features, partitioned into a 70% training set and a 30% held-out test set using the train_test_split function. A stratified split, implemented via the stratify parameter, was applied to preserve class balance across both subsets. To avoid data leakage and overly optimistic performance estimates, all preprocessing steps, including feature standardization using StandardScaler and Principal Component Analysis (PCA) from Scikit-learn, were encapsulated within a Pipeline and fitted exclusively on the training data within each cross-validation loop before being applied to the corresponding validation folds, as shown in Fig. 2.

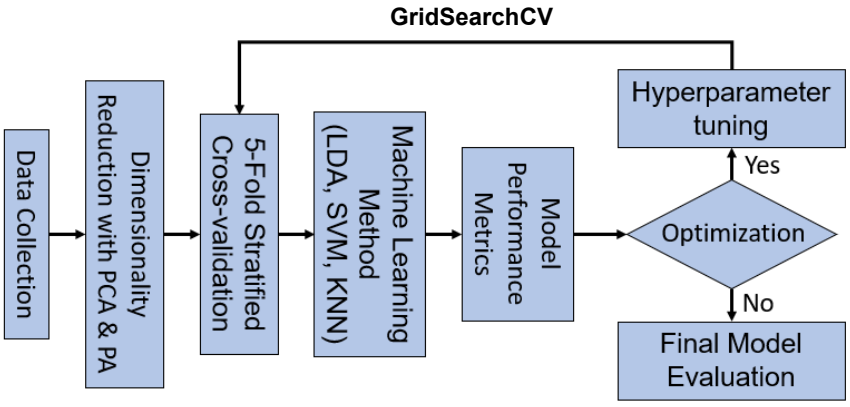

Figure 2. Workflow diagram of the proposed classification framework.

Principal component retention was determined using a combination of parallel analysis (PA) and a cumulative explained variance threshold of 95%. Parallel analysis compared eigenvalues from the training data with the 95th percentile of eigenvalues from randomly generated datasets using rng.normal noise. Based on this diagnostic analysis, the maximum value among these criteria was used, resulting in 12 principal components being retained for the final machine learning pipeline.

Classification performance of three machine learning algorithms was compared using 5-fold stratified cross-validation (StratifiedKFold), namely Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), and K-Neighbors Classifier (KNN). Baseline performance was evaluated using models instantiated with their default parameters, followed by hyperparameter optimization using GridSearchCV to identify the best-performing model

configuration. The optimized models were then evaluated on the held-out test dataset, and confusion matrices were generated. Fig. 2 illustrates the process flowchart.

The classification performance of the machine learning models implemented for the polymer sample identification process was quantitatively evaluated using a set of performance metrics. These metrics include identification accuracy, precision, recall, and the F1-score, which are commonly used to assess machine learning algorithms in chemometric and material identification tasks. The derivation of these metrics utilizes data from the four fundamental outcomes of the classification task: True Positive (TP) and True Negative (TN), representing correct recognition of target and non-target classes, respectively, alongside False Positive (FP) and False Negative (FN), which denote instances of incorrect recognition. Machine learning model accuracy, precision, recall, and the F1-score are calculated as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (4)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision+Recall} \quad (5)$$

3. RESULTS

3.1. Spectrometer Prototype Performance Discussion

The camera's maximum operating wavelength was approximately 1100 nm due to limitations of the silicon-based photovoltaic sensor in the modified webcam [9]. Spectrometer accuracy was verified by comparing the detected value to the actual value of characteristic mercury emission peaks (436 nm and 546 nm) emitted from a fluorescent lamp [10] using the diffraction grating equation Eq. (1). Wavelengths at the measured angles, namely 28° and 35° as shown in Table 1, were compared with the actual wavelength of characteristic mercury emission peaks. Deviations of 27.6 nm (5.1%) for the 546 nm peak and 33.5 nm (7.6%) for the 436 nm peak were observed, as shown in Table 1.

Table 1. Deviation between theoretical and experimental values

Measured angle of diffraction	Experimental Value	Actual value	Wavelength difference	Error
28°	469.5 nm	436 nm	33.5 nm	7.6%
35°	573.6 nm	546 nm	27.6 nm	5.1%

These deviations were attributed to the low-cost imaging system, including radial lens distortion [11] and surface non-planarity in the educational-grade transmission grating, as shown in Fig. 3. Fig. 3 visualizes the extent of this surface irregularity by displaying the diffraction grating's warp profile. This profile was qualitatively assessed by capturing the reflection of a digital grid from a computer screen off the grating's surface. A perfectly planar surface would reflect the grid in straight lines; the observed reflected grid curved in certain areas, indicating a warped surface and contributing to the observed deviations. The resulting accuracy was nevertheless sufficient for pattern-based polymer classification provided that the warp profile remains constant.

https://doi.org/10.31436/iiumej.v27i3.4388

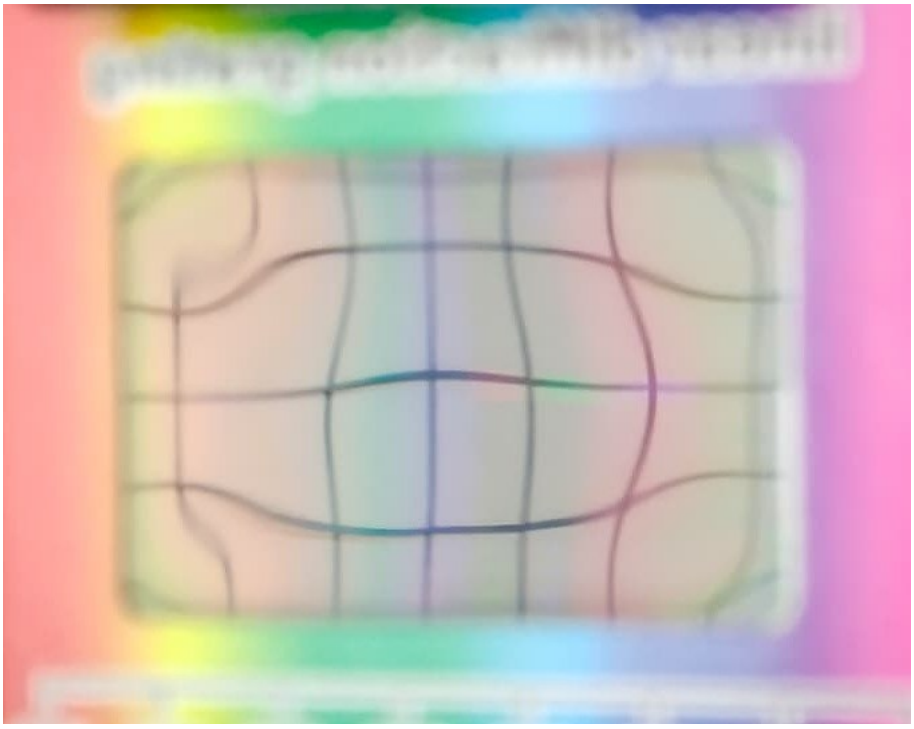

Figure 3. Reflection of grid with right angles across the diffraction grating.

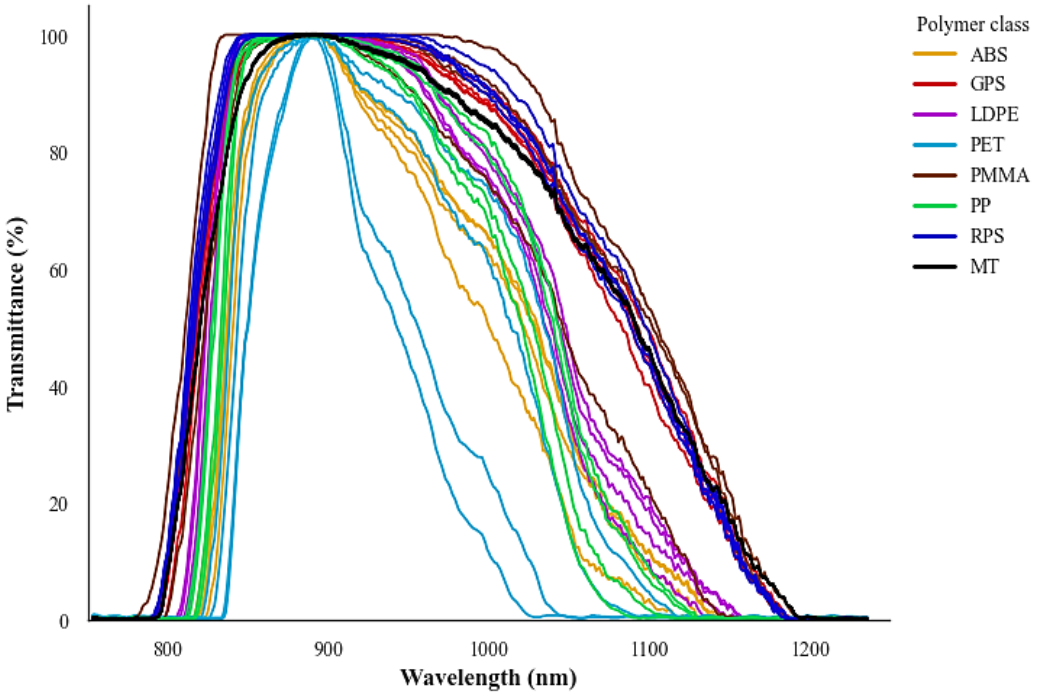

Figure 4. Different spectra of the seven lab-manufactured plastic samples and an empty baseline reference.

Although the low-cost optical components introduced minor deviations in absolute wavelength precision, classification performance was unaffected. The machine learning algorithms prioritize overall relative spectral features. Fig. 4 shows the different spectra of the seven lab-manufactured plastic samples and an empty baseline reference.

https://doi.org/10.31436/iiumej.v27i3.4388

Transparent samples have higher transmittance compared to semi-transparent samples. As a result, transparent samples such as GPS, RPS, and PMMA exhibit a wider range of high transmittance between 900 nm and 1200 nm, overlap with one another, and closely resemble the empty baseline reference (MT) spectrum.

3.2. Dimensionality Reduction Results Using PCA

Principal Component Analysis (PCA) was conducted using Parallel Analysis (PA), integrating a scree plot and a line indicating the 95th percentile eigenvalues obtained in Monte Carlo PA. Eigenvalues above this percentile indicate the optimal number of components to retain. As shown in Fig. 5, the PA results suggest a cutoff of 4 principal components (PCs) [12]. However, 12 PCs were required to explain 95% of the dataset's cumulative variance. Over-extraction is a less severe risk than under-extraction because the variation captured by successive components decreases sequentially [13]. In this context, retaining 12 components safeguarded against the negative effects of under-extraction. Therefore, 12 PCs were selected for subsequent machine learning development using SVM, LDA, and KNN classification models, which were commonly employed in plastic material identification [14]. For clarity, only the leading principal components are presented in the scree plot in Fig. 5.

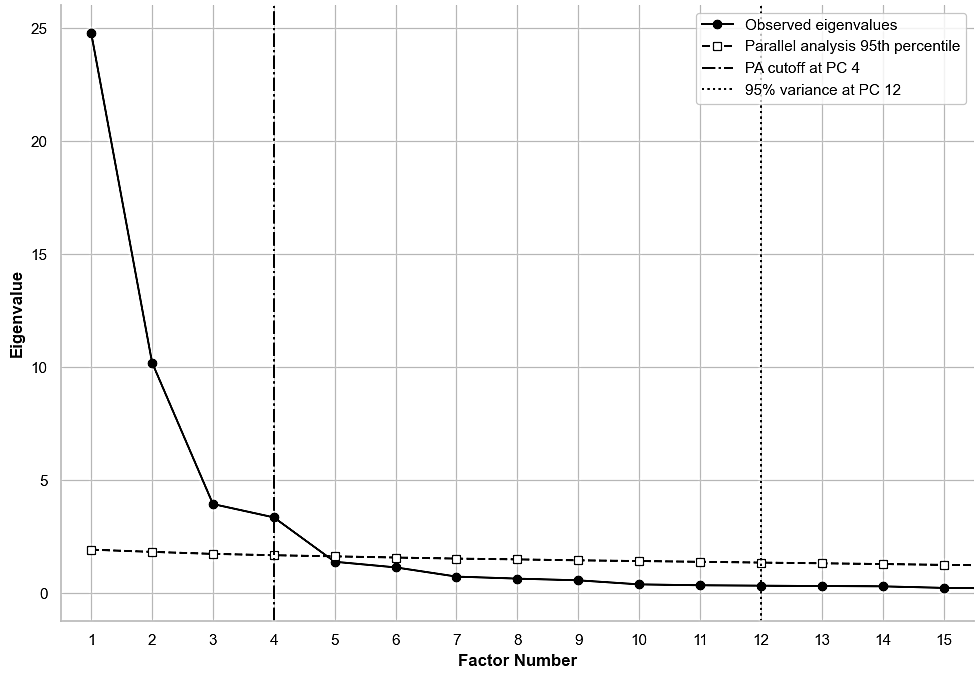

Figure 5. Parallel Analysis showing cut-off at 4 PCs

A 2D PCA projection, as shown in Fig. 6, capturing 69.88% of the variance from PC1 and PC2, was generated for visualization purposes to illustrate clustering patterns across classes. The empty baseline reference (MT) class displayed the tightest distribution. Similarly, Red Polystyrene (RPS) and Green Polystyrene (GPS) clustered very closely together. Meanwhile, PET exhibited the largest within-class spread, indicating higher intra-class variability. Partial class overlap was observed between the LDPE, PP, and ABS classes.

https://doi.org/10.31436/iiumej.v27i3.4388

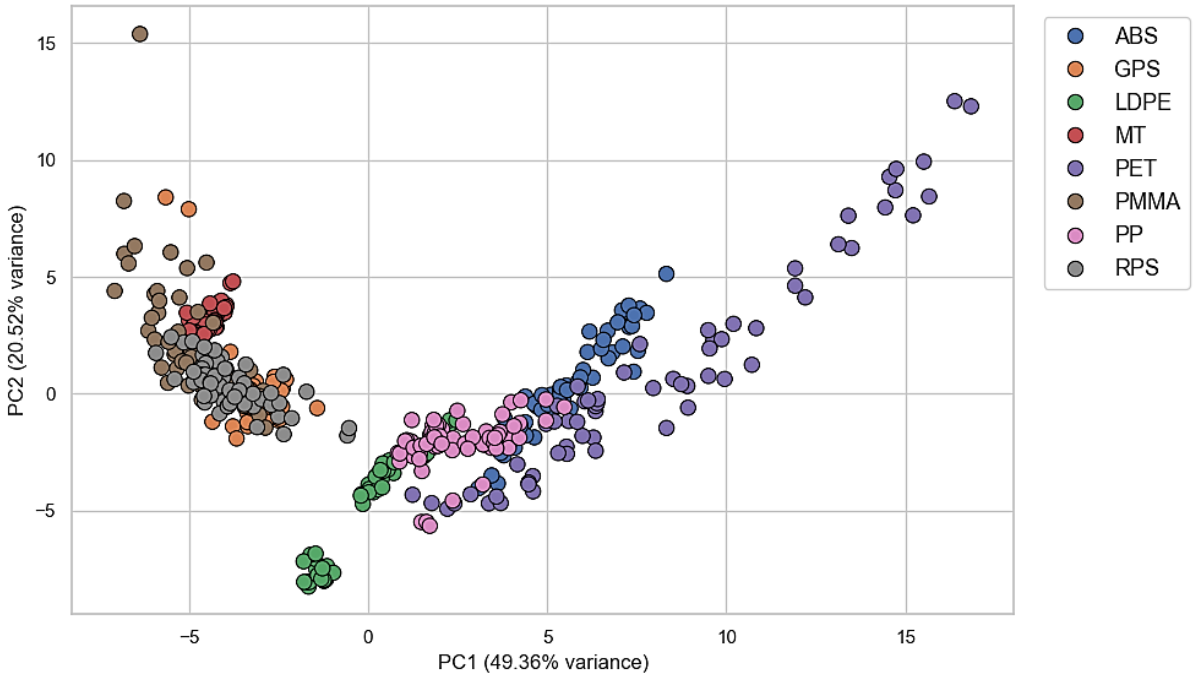

Figure 6. 2D PCA projection capturing 69.88% of the total variance.

3.3. Model Performance Results: Cross-Validation and Hyperparameter Tuning

Baseline classification performance was evaluated using 5-fold stratified cross-validation on the training set. The Support Vector Machine (SVM) achieved the highest baseline accuracy of 87.79%, followed by Linear Discriminant Analysis (LDA) at 84.66% and K-Nearest Neighbors (KNN) at 83.24%.

Following hyperparameter tuning, SVM and KNN achieved the highest held-out test accuracy, both reaching 82.24%. SVM used a linear kernel with a regularization parameter of C = 0.1, while KNN performed best with 7 neighbors and distance weighting. Although both models obtained the same test accuracy, SVM showed slightly better overall class-balanced performance, with a higher F1-score (82.36%) than KNN (81.98%), and was therefore selected as the preferred final model. In contrast, LDA showed lower held-out performance, with a test accuracy of 73.68%. Table 2 summarizes the results.

Table 2. Model performances following hyperparameter tuning

Model	PC (n)	Prediction (%)				
		Baseline Accuracy	Held-out test Accuracy	Precision	Recall	F-1 Score
SVM	**12**	**87.79**	**82.24**	**83.94**	**82.24**	**82.36**
LDA	12	84.66	73.68	75.32	73.68	73.71
KNN	12	83.24	82.24	83.94	82.24	81.98

3.4. Classification Behavior Across Prediction Classes

Table 3 further illustrates each prediction class's classification behavior in terms of precision, recall, and F1-score. LDA remains the weakest model, except that it has a higher precision score for the PMMA class than the KNN model. The SVM and KNN models showed varying strengths in predicting different classes. For example, the SVM model predicted ABS, PET, and PMMA better than KNN, while KNN performed better on GPS, LDPE, and PP.

https://doi.org/10.31436/iiumej.v27i3.4388

Nonetheless, all models classified the MT class perfectly, with precision, recall, and F1-score of 100%, confirming the consistency of the empty baseline reference spectrum.

For the SVM model, LDPE and RPS both showed low recall values of 63.16%, indicating that a substantial fraction of true samples from these classes were misclassified. This is understandable for the case of RPS due to its base polystyrene polymer, which is similar to that of GPS, as shown in the confusion matrix in Fig. 7. For the case of LDPE, the low recall value of 63.16%, indicating misclassification with the PP class as shown in Fig. 7, was largely due to the chemical similarity, being in the same polyolefin class [15] Furthermore, the SVM performance results highlight that PP exhibited a high recall of 89.47% but the lowest overall precision of 60.71%, meaning the SVM model also frequently overpredicted the PP class.

For KNN, RPS remained particularly problematic, with recall dropping even further to 52.63%. Like SVM, KNN also struggled with PP, showing high recall (89.47%) but relatively low precision (62.96%), further suggesting a tendency to overpredict the PP class.

Table 3. Precision, recall, and F1-score of the prediction classes by respective classification models.

Model	Metric	Scores (%)							
		ABS	GPS	LDPE	MT	PET	PMMA	PP	RPS
SVM	**Precision**	**100.00**	**75.00**	**85.71**	**100.00**	**94.12**	**80.95**	**60.71**	**75.00**
	Recall	**89.47**	**78.95**	**63.16**	**100.00**	**84.21**	**89.47**	**89.47**	**63.16**
	F1-score	**94.44**	**76.92**	**72.73**	**100.00**	**88.89**	**85.00**	**72.34**	**68.57**
LDA	Precision	80.95	62.50	66.67	100.00	100.00	76.47	56.00	60.00
	Recall	89.47	78.95	63.16	100.00	68.42	68.42	73.68	47.37
	F1-score	85.00	69.77	64.86	100.00	81.25	72.22	63.64	52.94
KNN	Precision	88.89	80.95	86.67	100.00	93.75	75.00	62.96	83.33
	Recall	84.21	89.47	68.42	100.00	78.95	94.74	89.47	52.63
	F1-score	86.49	85.00	76.47	100.00	85.71	83.72	73.91	64.52

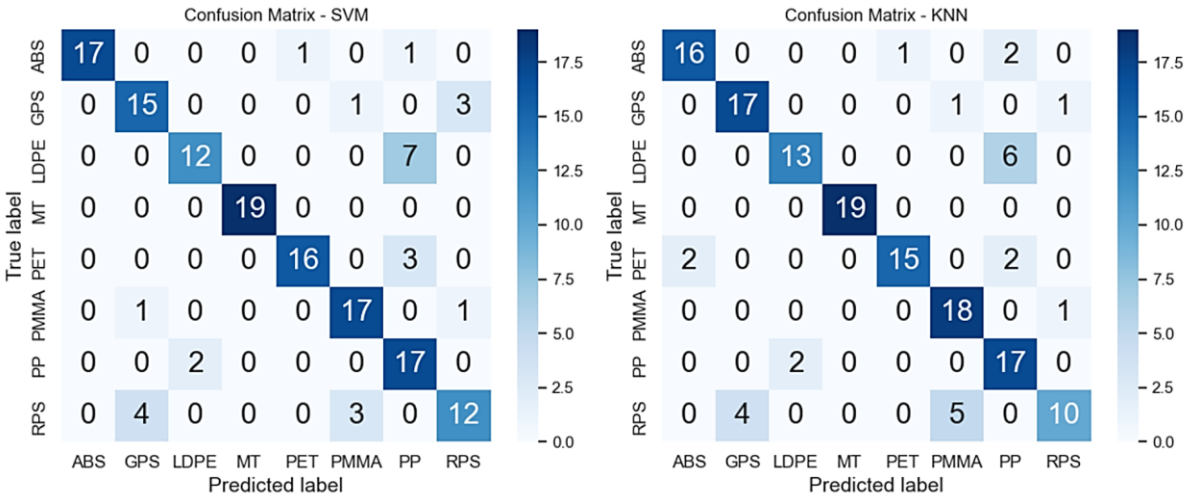

Figure 7. Confusion matrices of SVM and KNN from the held-out test set.

https://doi.org/10.31436/iiumej.v27i3.4388

4. DISCUSSION

4.1. Discussion of Model Performance and Polymer Identification

Previous research reported 72.50% accuracy for waste plastics when operating between 410 nm and 940 nm using Convolutional Neural Networks [16]. This demonstrated that low-cost near-infrared sensors combined with machine learning methods can recognize plastics effectively. This performance is comparable to the overall identification accuracy of 82.24% achieved in the current study. While high-end laboratory prototypes using Principal Component Analysis (PCA)-optimized filters have achieved accuracies ranging from 94.44% to 100%, those systems often utilize expensive InGaAs sensors to access wavelengths above 1100 nm [14]. The proposed prototype's efficacy suggests that continuous-spectrum acquisition, combined with robust chemometric models, can extract high-density information even from the restricted range of a silicon-based sensor.

4.2. Machine Learning Models Comparison

Selecting SVM, KNN, and LDA as classification models aligns with established benchmarks in the plastic recycling industry. These three algorithms are fundamental chemometric tools for handling the high dimensionality and overlapping spectral features of polymers [14].

In this study, Support Vector Machine (SVM) emerged as the most reliable classifier, achieving an F1-score of 82.36%. This performance is consistent with research showing that SVM offers superior classification efficacy for highly affordable short-wave near-infrared (SW-NIR) miniaturized spectrometers compared with Linear Discriminant Analysis (LDA) or K-Nearest Neighbors (KNN) [4]. These results also align with findings that identify SVM as an optimal model for rapid, on-site material classification using a handheld near-infrared (NIR) spectrometer [17].

SVM's superiority likely stems from its ability to use a linear kernel that models the linear relationship between absorbance and chemical composition described by the Beer-Lambert law while remaining robust to hardware-induced noise inherent in low-cost sensors. Furthermore, SVM is particularly effective at preventing underfitting and overfitting in high-dimensional spectral data, making it well suited for classification tasks with limited training samples [17].

4.3. Analysis of Polymer Feature Overlap

The observed confusion between LDPE and PP reflects a documented challenge in vibrational spectroscopy. Near-infrared spectroscopy often struggles to differentiate polyolefins such as LDPE and PP because of their extreme spectral similarity and chemical overlap. These polymers have nearly identical C-H stretching overtones—specifically the second overtone near 1200 nm—and differentiating them often requires high-precision instrumentation to resolve subtle differences in methyl group density and short-chain branching (SCB) [15].

Similarly, the spectral overlap between green and red polystyrene (GPS and RPS) is anticipated, as both variants share an identical molecular fingerprint associated with the polystyrene base polymer. While specific colorants can shift absolute transmittance, these additives do not fundamentally alter the polymer backbone's molecular vibrations.

4.4. Study Limitations and Socio-Economic Implications

The primary physical limitation of this study is the silicon bandgap of 1.12 eV, which restricts the prototype's detection capability to wavelengths below 1100 nm [9]. Beyond this threshold, standard silicon sensors become "blind" to the more prominent second overtones of C-H vibrations found in industrial NIR applications [14].

Despite these technical trade-offs, the system provides a necessary technological bridge for decentralized recycling. Industrial units like Titech's PolySort can process up to 5 tons per hour [18], but their high investment costs, such as hyperspectral imaging (HSI) technology of up to $ 50,000, are prohibitive for community workshops [5]. The proposed 3D-printed modular framework reduces these costs to a few hundred dollars, democratizing access to scientifically sound material segregation tools for fablabs and small-scale recovery initiatives.

5. CONCLUSION

This study successfully demonstrated the development of a modular, cost-effective, and 3D-printed transmission NIR spectrometer capable of operating in the 850 nm to 1100 nm range using open-source tools and commercially available optical components. By integrating machine learning with Principal Component Analysis (PCA) for dimensionality reduction, the system effectively extracted classification patterns from a restricted spectral range. Among the evaluated models, both the Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) achieved a held-out test accuracy of 82.24%, with SVM proving to be the most reliable overall classifier due to its superior class-balanced performance (F1-score of 82.36%).

While the hardware setup introduced wavelength deviations, specifically a 7.6% error at the 436 nm peak and 5.1% at the 546 nm peak, these were attributed to inherent optical limitations such as radial lens distortion and surface non-planarity in the educational-grade transmission diffraction grating. Despite these absolute precision errors, the machine learning algorithms maintained high performance by prioritizing relative spectral features, successfully identifying the empty baseline (MT) with 100% accuracy. However, challenges remain in distinguishing polymers with significant chemical similarity or feature overlap, such as LDPE with PP and RPS with GPS.

In conclusion, this research provides a viable framework for creating low-cost identification tools that bridge the gap between expensive laboratory-grade instruments and the practical needs of decentralized, community-level plastic segregation. Such accessible technology is essential to reduce contamination in mixed plastic waste streams and accelerate the transition toward a sustainable circular economy.

ACKNOWLEDGMENT

The authors acknowledge the Ministry of Higher Education, Malaysia, for the financial support provided through the Fundamental Research Grant Scheme (Ref. No. FRGS/1/2022/TK09/USM/01/4).

REFERENCES

[1] Macheca AD, Mutuma B, Adalima JL, Midheme E, Lúcas LHM, Ochanda VK, Mhlanga SD. (2024) Perspectives on plastic waste management: challenges and possible solutions to ensure its sustainable use. Recycling, 9(5): 77. https://doi.org/10.3390/recycling9050077

[2] Lahl U, Zeschmar-Lahl B. (2024) Material recycling of plastics—a challenge for sustainability. Sustainability, 16(15): 6630. https://doi.org/10.3390/su16156630

[3] Bernat K. (2023) Post-consumer plastic waste management: from collection and sortation to mechanical recycling. Energies, 16(8): 3504. https://doi.org/10.3390/en16083504

[4] Grabska J, Huck CW, Bec KB. (2021) Principles and applications of miniaturized near-infrared (NIR) spectrometers. Chem. Eur. J., 27(5): 1514–1532. https://doi.org/10.1002/chem.202002838

[5] Werner T, Dawoud M, Aschenbrenner D, Taha I. (2025) Cost-efficient detection of plastics from post-consumer packaging waste using selected bands in the near-infrared spectrum. Macromol. Mater. Eng., 2025: 2500143. https://doi.org/10.1002/mame.202500143

[6] Silva WRF, Suarez WT, Reis C, dos Santos VB, Carvalho FA, Reis EL, Vicentini FC. (2021) Multifunctional webcam spectrophotometer for performing analytical determination and measurements of emission, absorption, and fluorescence spectra. J. Chem. Educ., 98(4): 1442–1447. https://doi.org/10.1021/acs.jchemed.0c01085

[7] Straller A, Gessler B. (2019) Identification of plastic types using discrete near infrared reflectance spectroscopy. ResearchGate. https://www.researchgate.net/publication/337868860_Identification_of_Plastic_Types_Using_Discrete_Near_Infrared_Reflectance_Spectroscopy

[8] Laganovska K et al. (2020) Portable low-cost open-source wireless spectrophotometer for fast and reliable measurements. HardwareX, 7: e00108. https://doi.org/10.1016/j.ohx.2020.e00108

[9] Huang S et al. (2024) Broadband-spectral-responsivity of black silicon photodetector with high gain and sub-bandgap sensitivity by titanium hyperdoping. Opt. Laser Technol., 171: 110399. https://doi.org/10.1016/j.optlastec.2023.110399

[10] Rodrigues M, Marques MB, Carvalho PS. (2016) How to build a low cost spectrometer with Tracker for teaching light spectra. Phys. Educ., 51(1): 014002. https://doi.org/10.1088/0031-9120/51/1/014002

[11] Liebold F, Mader D, Sardemann H, Eltner A, Maas HG. (2023) A bi-radial model for lens distortion correction of low-cost UAV cameras. Remote Sens., 15(22): 5283. https://doi.org/10.3390/rs15225283

[12] Ledesma RD, Valero-Mora P, Macbeth G. (2015) The scree test and the number of factors: a dynamic graphics approach. Span. J. Psychol., 18: E11. https://doi.org/10.1017/sjp.2015.13

[13] Peres-Neto PR, Jackson DA, Somers KM. (2005) How many principal components? stopping rules for determining the number of non-trivial axes revisited. Comput. Stat. Data Anal., 49(4): 974–997. https://doi.org/10.1016/j.csda.2004.06.015

[14] Neo ERK, Yeo Z, Low JSC, Goodship V, Debattista K. (2022) A review on chemometric techniques with infrared, Raman and laser-induced breakdown spectroscopy for sorting plastic waste in the recycling industry. Resour. Conserv. Recycl., 180: 106217. https://doi.org/10.1016/j.resconrec.2022.106217

[15] Li S, Ivancic RJS, Sutliff BP, Huang D, Blázquez-Blázquez E, Martin TB, Migler KB, Audus DJ, Orski SV. (2026) Predicting properties from near-infrared spectra with machine learning for improved polyolefin differentiation. ACS Polym. Au, 6(4): 246–255. https://doi.org/10.1021/acspolymersau.5c00131

[16] Martinez-Hernandez U, West G, Assaf T. (2024) Low-cost recognition of plastic waste using deep learning and a multi-spectral near-infrared sensor. Sensors, 24(9): 2821. https://doi.org/10.3390/s24092821

[17] Yang Y, Zhang X, Yin J, Yu X. (2020) Rapid and nondestructive on-site classification method for consumer-grade plastics based on portable NIR spectrometer and machine learning. J. Spectrosc., 2020: 6631234. https://doi.org/10.1155/2020/6631234

[18] Brunner S, Fomin P, Kargel C. (2015) Automated sorting of polymer flakes: fluorescence labeling and development of a measurement system prototype. Waste Manag., 49: 600-607. https://doi.org/10.1016/j.wasman.2014.12.006